A note on the estimation of stochastic and deterministic production frontiers with maximum entropy

Pedro Macedo

CIDMA - Center for Research & Development in Mathematics and Applications, Department of Mathematics, University of Aveiro, 3810-193 Aveiro, Portugal, pmacedo@ua.pt

ABSTRACT

Stochastic frontier analysis with maximum entropy estimation has received considerable attention in the literature in the recent years. In this work, the estimation of stochastic and deterministic production frontiers with maximum entropy methods, the advantages and disadvantages relatively to maximum likelihood, and some proposals to improve maximum entropy estimation in this context are presented and discussed.

Keywords: *Efficiency analysis, Ill-posed models, Info-metrics*

1. INTRODUCTION

Aigner et al. (1977), Battese and Corra (1977) and Meeusen and van Den Broeck (1977) were the pioneers of stochastic frontier analysis (SFA). A general stochastic frontier model can be represented as

$$\ln y_n = f(\boldsymbol{x}_n, \boldsymbol{\beta}) + v_n - u_n, \tag{1}$$

where y_n is the scalar output for producer n $(1, 2, \dots, N)$, $f(.)$ represents the production frontier, $\boldsymbol{x}_n$ is a row vector with logarithms of inputs, $\boldsymbol{\beta}$ is a column vector of parameters, v is a noise component (measurement errors, random shocks, etc.) and $u \geq 0$ is a one-sided component representing technical inefficiency. If the noise term v in model (1) is removed, a deterministic production frontier (DPF) model is obtained.

The parameters of model (1) are usually estimated through maximum likelihood (ML). The random variable v is usually assumed to be normally distributed, $N(0, \sigma_v^2)$, and the u is defined through different distributions, such as exponential, nonnegative half normal, truncated normal or gamma; e.g., Kumbhakar and Knox Lovell (2000). It is important to note that these distributional assumptions are the main criticism on SFA, in particular the choice of the distribution for the u error component, since different distributional assumptions can lead to different predictions of technical efficiency.

An interesting alternative to ML is maximum entropy (ME) estimation. Golan et al. (1996) developed the generalized maximum entropy (GME) and the generalized cross entropy (GCE) estimators, which can be used in ill-posed models (e.g., models affected by collinearity, under-determined models, micronumerosity, non-normal errors). Recently, due to the fact that frontier models are usually ill-posed, an increasing interest with GME and GCE in SFA has emerged in the literature; e.g., Campbell et al. (2008), Rezek et al. (2011), Macedo and Scotto (2014), Robaina-Alves et al. (2015) and Moutinho et al. (2018a).

2. MAXIMUM ENTROPY ESTIMATION IN SFA AND DPF MODELS

Considering the general stochastic frontier model in (1) defined in matricial form as

$$\ln \boldsymbol{y} = f(\boldsymbol{X}, \boldsymbol{\beta}) + \boldsymbol{v} - \boldsymbol{u}, \tag{2}$$

the reparameterization of the $(K \times 1)$ vector $\boldsymbol{\beta}$ and the $(N \times 1)$ vector $\boldsymbol{v}$ follows the same procedures as in the traditional regression model with GME and GCE estimators; e.g., Golan et al. (1996) and Golan (2018). Each parameter is treated as a discrete random variable with a compact support and $M \geq 2$ possible outcomes, and each error is defined as a finite and discrete random variable with $J \geq 2$ possible outcomes. The reparameterizations are given by $\boldsymbol{\beta} = \boldsymbol{Z}\boldsymbol{p}$, where $\boldsymbol{Z}$ is a $(K \times KM)$ matrix of support points and $\boldsymbol{p}$ is a $(KM \times 1)$ vector of unknown probabilities to be estimated, and $\boldsymbol{v} = \boldsymbol{A}\boldsymbol{w}$, where $\boldsymbol{A}$ is a $(N \times NJ)$ matrix of support points and $\boldsymbol{w}$ is a $(NJ \times 1)$ vector of unknown probabilities to be estimated. Extending this idea to the vector $\boldsymbol{u}$, the reparameterization is similar to the one conducted for the random variable representing noise, $\boldsymbol{v}$, taking only into account that $\boldsymbol{u}$ is a one-sided random variable, which implies that the lower bound for the supports (with $L \geq 2$ points) is zero for all error values. The reparameterization of $\boldsymbol{u}$ can be defined by $\boldsymbol{u} = \boldsymbol{B}\boldsymbol{\rho}$, where $\boldsymbol{B}$ is a $(N \times NL)$ matrix of support points and $\boldsymbol{\rho}$ is a $(NL \times 1)$ vector of unknown probabilities to be estimated.

Thus, the GME estimator in the SFA context can be defined by

$$\operatorname{argmax}_{\boldsymbol{p},\boldsymbol{w},\boldsymbol{\rho}} \left\{ -(1-\theta)\boldsymbol{p}' \ln \boldsymbol{p} - \frac{\boldsymbol{\theta}}{\mathbf{2}} \boldsymbol{w}' \ln \boldsymbol{w} - \frac{\theta}{2} \boldsymbol{\rho}' \ln \boldsymbol{\rho} \right\}, \tag{3}$$

subject to the model constraints,

$$\ln \boldsymbol{y} = \boldsymbol{XZp} + \boldsymbol{Aw} - \boldsymbol{B\rho}, \tag{4}$$

and the additivity constraints,

$$\begin{aligned} \mathbf{1}_K &= (\boldsymbol{I}_K \otimes \mathbf{1}_M')\boldsymbol{p}, \\ \mathbf{1}_N &= \left(\boldsymbol{I}_N \otimes \mathbf{1}_J'\right)\boldsymbol{w}, \\ \mathbf{1}_N &= (\boldsymbol{I}_N \otimes \mathbf{1}_L')\boldsymbol{\rho}, \end{aligned} \tag{5}$$

where $\otimes$ represents the Kronecker product and $\theta \in (0, 1)$ assigns different weights on the components of the objective function (by default, $\theta = 2/3$; the value can be obtained by some kind of rotational estimation through the minimization of a given loss function). Accordingly, the GME estimator in the DPF model context can be defined by

$$\operatorname{argmax}_{\boldsymbol{p},\boldsymbol{\rho}}\{-(1-\theta)\boldsymbol{p}' \ln \boldsymbol{p} - \theta\boldsymbol{\rho}' \ln \boldsymbol{\rho}\}, \tag{6}$$

subject to the model constraints,

$$\ln \boldsymbol{y} = \boldsymbol{XZp} - \boldsymbol{B\rho}, \tag{7}$$

and the additivity constraints,

$$\begin{aligned} \mathbf{1}_K &= (\boldsymbol{I}_K \otimes \mathbf{1}_M')\boldsymbol{p}, \\ \mathbf{1}_N &= (\boldsymbol{I}_N \otimes \mathbf{1}_L')\boldsymbol{\rho}, \end{aligned} \tag{8}$$

where $\theta \in (0, 1)$ assigns different weights on the components of the objective function ($\theta = 1/2$, by default; as previously, the value can be obtained by cross-validation).

On the other hand, the GCE formulation in SFA can be defined by

$$\operatorname{argmin}_{\boldsymbol{p},\boldsymbol{w},\boldsymbol{\rho}}\left\{(1-\theta)\boldsymbol{p}' \ln\left(\frac{\boldsymbol{p}}{\boldsymbol{q}_1}\right) + \frac{\boldsymbol{\theta}}{\mathbf{2}}\boldsymbol{w}' \ln\left(\frac{\boldsymbol{w}}{\boldsymbol{q}_2}\right) + \frac{\theta}{2}\boldsymbol{\rho}' \ln\left(\frac{\boldsymbol{\rho}}{\boldsymbol{q}_3}\right)\right\}, \tag{9}$$

subject to the model and the additivity constraints (4) and (5), respectively, and $\theta = 2/3$, by default. The vectors $\boldsymbol{q}_i$ $(i = 1, 2, 3)$ represent prior information; e.g., Macedo et al. (2014). The GCE formulation in DPF models is easily defined accordingly. Finally, the GME and GCE estimators have no closed form solutions, which mean that numerical optimization techniques are required to solve these statistical problems.

The support matrices $\boldsymbol{Z}$ and $\boldsymbol{A}$ are defined by the researcher based on prior information; see Macedo and Scotto (2014) and Henderson et al. (2015) for further details. To discuss the definition of matrix $\boldsymbol{B}$, it is important to note that the traditional distributional assumptions concerning the error inefficiency component have been used in empirical work since it is expected a particular behavior in the distribution of technical inefficiency predictions. For example, in the discussion of the normal – half normal model, a popular one in empirical work, Kumbhakar and Knox Lovell (2000: 74) argued that the choice of the latter distribution "is based on the plausible proposition that the modal value of technical inefficiency is zero, with increasing values of technical inefficiency becoming increasingly less likely."

Thus, following this reasoning, an important advantage of the GME and GCE estimators is that distributional assumptions for the two-error component are not necessary, but the same beliefs can be expressed in the models through the error supports (in GME) or through the vectors with prior information (in GCE). For example, to define supports in matrix $\boldsymbol{B}$ within the GME estimator, Campbell et al. (2008) suggested the use of the mean of the data envelopment analysis (DEA) and SFA efficiency predictions to define the supports with a specific upper bound (ub) defined as

$$\boldsymbol{b}_n' = [0, 0.005, 0.01, 0.015, \mathrm{ub}], \tag{10}$$

considering five points, and Macedo et al. (2014) suggested supports defined as

$$\boldsymbol{b}_n' = [0, 0.01, 0.02, 0.03, -\ln(\mathrm{DEA}_n)], \tag{11}$$

where DEA_n can represent the lower technical efficiency prediction obtained by DEA in the N observations of the sample, and also considering five points in the supports. Naturally, other information can be used to define the upper bound. For example, considering an (almost totally) inefficient producer, note that can be used: $-\ln(\mathrm{DEA}_n) = -\ln(0.0067) \approx 5$. Although the definition of this prior information deserves future research, some recent results suggest that the rankings of efficiency predictions are not very sensitive to the definition of this upper bound; e.g., Moutinho et al. (2018b).

For the GCE estimator, since only the vector $\boldsymbol{q}_3$ is non-uniform following the prior beliefs mentioned previously, considering five points it can follow the structure

$$\boldsymbol{q}_3 = [0.40, 0.30, 0.15, 0.10, 0.05]', \tag{12}$$

or a similar one, for each observation, where the cross-entropy objective shrinks the posterior distribution in order to have more mass near zero. With GCE estimation, the supports in matrix $\boldsymbol{B}$ can be defined with five equally spaced points in the interval $[0, -\ln(\mathrm{DEA}_n)]$, where DEA_n can represent the technical efficiency prediction obtained by DEA in the N observations of the sample; see Macedo and Scotto (2014) for further details. Again, of course, other information can be used to define this upper bound (see the example above).

Table 1: Advantages and disadvantages of ML and ME estimation in DPF/SFA

	ML estimation	ME estimation
Advantages	• theory is well established; • empirical work is massive in the literature.	• distributional assumptions for the two-error component are not required; • available for well- and ill-posed models, which means that desirable sophisticated production frontiers can be used (e.g., translog family).
Disadvantages	• distributional assumptions for the two-error component are required; • ML estimator may not be unique and it is attractive mainly due to its large-sample properties; • possible convergence problems; e.g., Meesters (2014); • only available for well-posed models.	• possible convergence problems; • error supports for GME or vectors with prior weights for GCE are needed.

The main advantages and disadvantages of both estimation procedures in the SFA context are briefly presented in Table 1. Comparing the main characteristics of both procedures, the use (or not) of distributional assumptions for the inefficiency error component play a central role between them. In fact, in ME estimation, the choice of the three central values in (10) and (11), that define the prior mean and the skewness within the GME estimator, or the vector $\boldsymbol{q}_3$ within the GCE estimator, is the main difficulty that can discourage the use of ME estimators.

3. HOW TO DEFINE PRIOR INFORMATION FOR THE INEFFICIENCY ERROR COMPONENT?

A possible solution to define prior information needed in ME estimation, not considered so far, is the use of the skewness value of ordinary least squares (OLS) residuals. It is well-known in efficiency analysis literature that OLS residuals can be used to test the presence of technical inefficiency; e.g., Kumbhakar and Knox Lovell (2000: 73). Since the random variable v in model (1) is assumed to be symmetrically distributed, the skewness of the OLS residuals represents the skewness of the one-sided random variable of technical inefficiency.

Note that if $u > 0$, then the composed error is negatively skewed, which suggests the presence of technical inefficiency.[1] Thus, in empirical work, given the skewness of the OLS residuals, even in the case of "wrong" skewness, the supports in (10) and (11) can be defined in order to reflect this information.

However, if the model is ill-posed, the information from OLS may not be available, as well as the mean of efficiency predictions from ML. How to define prior information in such cases? One possible solution is the use of the skewness value of GME/GCE residuals, in the same sense to the use of the OLS residuals, using minimal prior information (support spaces, centered on zero, with large amplitude) on GME/GCE estimators.[2] Another possible solution is based on the moment generating function of the truncated normal distribution. Meesters (2014) shows that the three most commonly used distributions in SFA (half normal, truncated normal and exponential) may be all represented by the truncated normal distribution.[3] This is an important finding because it guides researchers' attention to the truncated normal distribution and the normal – truncated normal model. Thus, the moment generating function of the truncated normal distribution, namely the first and the third moments, may be used to define the prior information needed in ME estimation.

Naturally, these proposed approaches only provide guidelines for the prior information required, since there are different possibilities to define supports based on the information from the residuals of the OLS estimation or from the moment generating function of the truncated normal distribution. However, as mentioned by Rezek et al. (2011: 364), it is important to note that the selection "of these vectors sets a prior expectation of mean efficiency; however, it does not preordain that result." Additionally, since incorrect prior information does not constrain the solution if it is not consistent with the data (Golan et al., 1996), it is expected that the GCE estimator remains stable when the vector $\boldsymbol{q}_3$ comprises possible incorrect prior information.

---

[1] The concept of "full efficient producer" is fundamental, but it is sometimes wrongly interpreted, because there is always inefficiency in the production activity. Moreover, the presence of technical inefficiency is evaluated by hypothesis testing and, in the majority of the cases, even without any statistical test, it is known that the null hypothesis is false; e.g., it will be a parameter *exactly equal* to the test value? This wrong interpretation can lead to inappropriate considerations about the existence of technical efficiency.

[2] Additionally, the normalized entropy for the composed error structure provides important information on the noise component, namely the possible existence of technical inefficiency and the distribution of mass in supports.

[3] The paper also alerts for possible convergence problems with ML estimation, an issue that is usually ignored in practice.

4. CONCLUDING REMARKS

Maximum entropy estimators (information-theoretical methods) appear to be powerful alternatives to traditional ML estimation in DPF and SFA. Some advantages of ME estimators in this context are: (a) the possibility of considering prior information on the parameters and errors' components; (b) the traditional assumptions on the errors' distributions (half normal, truncated normal, exponential, among many others) are not necessary; (c) they can be used in ill-posed production frontier models, such as under-determined models (e.g., when are used sophisticated production functions) or in models affected by severe collinearity, which restrict the use of traditional estimators (ML or corrected ordinary least squares estimators); and (d) they can be used in different DPF/SFA structures (state-contingent, cross-sectional, panel data, two-tier, among many others, including frontier models with heteroscedasticity).

With the previous proposals to define supports in matrix $\boldsymbol{B}$, the DEA is used in SFA with ME estimation only to define an upper bound for the supports, which means that the main criticism on DEA is used here as an advantage. On the other hand, the main criticism on SFA with ML is avoided with ME estimation, because the composed error structure is used without specific statistical distributional assumptions. Although in this work only the GME and GCE estimators are discussed, other information-theoretical methods are easily adapted for the DPF and SFA contexts (e.g., the GME-α estimators, with Rényi and Tsallis entropies).

An important issue that will deserve further investigation is a sensitivity analysis on the efficiency predictions given the prior information used in ME estimation, in the same reasoning that Kumbhakar and Knox Lovell (2000: 90) answering to the question "Do Distributional Assumptions Matter?" with ML estimation, argued that "Sample mean efficiencies are no doubt apt to be sensitive to the distribution assigned to the one-sided error component (…) What is not so clear is whether a ranking of producers by their individual efficiency scores, or the composition of the top and bottom efficiency scores deciles, is sensitive to distributional assumptions." Recent results in Moutinho et al. (2018b) suggest that the rankings of efficiency are not very sensitive to the prior information used by ME methods, but extensive simulation studies are needed in future research to validate this assertion.

Acknowledgments

The author was supported by Fundação para a Ciência e a Tecnologia (FCT), within project UID/MAT/04106/2019 (CIDMA).

References

Aigner, D., Knox Lovell, C. and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics,* 6(1), 21-37.

Battese, G. and Corra, G. (1977). Estimation of a production frontier model: with application to the pastoral zone of Eastern Australia. *The Australian Journal of Agricultural and Resource Economics*, 21(3), 169-179.

Campbell, R., Rogers, K. and Rezek, J. (2008). Efficient frontier estimation: a maximum entropy approach. *Journal of Productivity Analysis,* 30, 213-221.

Golan, A. (2018). *Foundations of Info-Metrics: Modeling, Inference, and Imperfect Information*. New York, Oxford University Press.

Golan, A., Judge, G. and Miller, D. (1996). *Maximum Entropy Econometrics: Robust Estimation with Limited Data*. Chichester, John Wiley & Sons.

Henderson, H., Golan, A. and Seabold, S. (2015). Incorporating prior information when true priors are unknown: An Information-Theoretic approach for increasing efficiency in estimation. *Economics Letters*, 127, 1-5.

Kumbhakar, S. C. and Knox Lovell, C. (2000). *Stochastic Frontier Analysis*. Cambridge, Cambridge University Press.

Macedo, P. and Scotto, M. (2014). Cross-entropy estimation in technical efficiency analysis. *Journal of Mathematical Economics*, 54, 124-130.

Macedo, P., Silva, E. and Scotto, M. (2014). Technical efficiency with state-contingent production frontiers using maximum entropy estimators. *Journal of Productivity Analysis,* 41(1), 131-140.

Meesters, A. (2014). A note on the assumed distributions in stochastic frontier models. *Journal of Productivity Analysis,* 42(2), 171-173.

Meeusen, W. and van Den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review*, 18(2), 435-444.

Moutinho, V., Madaleno, M., Macedo, P., Robaina, M. and Marques, C. (2018a). Efficiency in the European agricultural sector: environment and resources. *Environmental Science and Pollution Research*, 25(18), 17927-17941.

Moutinho, V., Robaina, M. and Macedo, P. (2018b). Economic-environmental efficiency of European agriculture – a generalized maximum entropy approach. *Agricultural Economics – Czech*, 64(10), 423-435.

Rezek, J., Campbell, R. and Rogers, K. (2011). Assessing Total Factor Productivity Growth in Sub-Saharan African Agriculture. *Journal of Agricultural Economics*, 62(2), 357-374.

Robaina-Alves, M., Moutinho, V. and Macedo, P. (2015). A new frontier approach to model the eco-efficiency in European countries. *Journal of Cleaner Production*, 103, 562-573.